DOCUMENT RESUME

ED 376 195 TM 022 293

AUTHOR Hipps, Jerome A.
TITLE Trustworthiness and Authenticity: Alternate Ways To Judge Authentic Assessments.
PUB DATE Apr 93
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Constructivism (Learning); *Credibility; *Educational Assessment; Educational Philosophy; Elementary Secondary Education; *Evaluation Methods; Models; Reliability; Student Evaluation; *Validity
IDENTIFIERS Alternative Assessment; *Authentic Assessment; Authenticity; Guba (Egon); Lincoln (Yvonna S); Objectivity; Stakeholder Evaluation

ABSTRACT

New methods are needed to judge the quality of alternative student assessment, methods which complement the philosophy underlying authentic assessments. This paper examines assumptions underlying validity, reliability, and objectivity, and why they are not matched to authentic assessment, concentrating on the constructivist paradigm of E. Guba and Y. Lincoln and its synergy with authentic assessment. Guba and Lincoln have rejected validity, reliability, generalizability, and objectivity as criteria for judging inquiries conducted within the constructivist paradigm. They have suggested trustworthiness and authenticity as frameworks for judging the quality of studies. Trustworthiness consists of four elements: credibility, transferability, dependability, and confirmability. These elements parallel validity, generalizability, reliability, and objectivity. Authenticity criteria, which focus on knowing, action, and fairness, have no counterparts in the conventional research paradigm, but are primarily demonstrated through stakeholder testimony and are supported by an audit trail of evidence of fairness and authenticity. The constructivist paradigm offers new lenses for judging the quality of authentic assessments. (Contains 12 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 376 195

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
JEROME A. HIPPS
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

Trustworthiness and Authenticity: Alternate Ways to Judge Authentic Assessments

Jerome A. Hipps

San Francisco Unified School District

Paper to be presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA, April, 1993.

Jerome A. Hipps
Testing, Evaluation & Research
San Francisco Unified School District
135 Van Ness Avenue, Room 314
San Francisco, CA 94102
(415) 241-6400

TM022293

BEST COPY AVAILABLE

Trustworthiness and Authenticity:
Alternate Ways to Judge Authentic Assessments

Jerome A. Hipps

New, alternative modes of assessing student academic performance have become increasingly popular in the past several years. Proponents of these new modes of assessment, which include portfolios and performance based assessments, believe they provide a more accurate and complete picture of a student's abilities than traditional paper-and-pencil assessment techniques (Wolf, Bixby, Glen, & Gardner, 1991). The rise in the popularity of alternative assessments has been accompanied by concerns about their quality. Are alternative assessments valid? Are they reliable? These questions come from traditional measurement theory which has its roots in logical positivism. However, they are not appropriate questions to ask of authentic assessments. At their core, authentic assessments do not grow out of logical positivism. They are built on different assumptions about students and how to collect data about what they know and can do. As a result, different approaches are needed to judge the quality of authentic assessments, approaches which are in keeping with the philosophy underlying authentic assessments.

This paper examines the assumptions underlying validity, reliability, and objectivity and why they are not matched to authentic assessment. It proceeds to discuss Guba and Lincoln's constructivist paradigm and its synergy with authentic assessments (Guba & Lincoln, 1989; Lincoln & Guba, 1985). Guba and Lincoln have rejected validity, reliability, generalizabilty, and objectivity as criteria that are applicable when judging inquiries conducted within the constructivist paradigm. They have suggested using two other frameworks, trustworthiness and authenticity, when judging the quality of studies. After discussing trustworthiness and authenticity, this paper examines how these frameworks might be applied to authentic assessments to support high quality assessments.

What's Wrong With Validity, Reliability, and Objectivity?

Reliability, validity, and objectivity are key criteria that are used when judging the quality of measures of traits and abilities. These concepts are rooted in traditional measurement theory which is based on logical positivism. There are several assumptions underlying the use of these criteria which deserve examination. One assumption is that the quality being measured is fixed, such as weight or height. An example is how a construct like intelligence is measured. Tests like the Stanford-Binet assume two things. First, intelligence is seen as a fixed characteristic. Second, the score provided by an intelligence test is assumed to be an accurate measure of that characteristic. Standardized achievement tests are built on similar assumptions. For example, a student's skill in reading comprehension is treated as being at a fixed level which is then measured by the reading comprehension subtest.

A second assumption underlying measurement theory involves the relationship between the inquirer and the object of measurement. Since the ability or quality being measured is seen as a fixed property of the person, the goal becomes measuring that quality *as it is*. One does this procedurally using "objective" measures to minimize the chance that the inquirer or the inquiry process affect what is actually being measured. In practical terms with standardized tests, objectivity is achieved by administering paper and pencil tests to students under uniform test conditions. Uniform conditions are achieved by doing such things as reading a standard set of instructions to students and allowing them a fixed amount of time to complete the test. Each student is treated in the same fashion.

Authentic assessments challenge the assumptions underlying validity, reliability, and objectivity. First, advocates of authentic assessments see achievement in a given subject area as multi-dimensional. Its multi-dimensional nature makes it difficult to reduce achievement to a single number. Authentic assessments seek to provide broader, holistic views of students. This holistic orientation contrasts the traditional approach where what students know is frequently measured in terms of many discrete objectives and variables.

Authentic assessments also challenge traditional measurement theory's demand for objectivity. A hallmark of authentic assessments is the opportunities they provide for students and teachers to interact during the assessment process. This interaction is

beneficial and desirable for at least two reasons. First, the teacher is able to probe the student to clarify what a student may be saying. This helps the teacher to better understand the boundaries of what the student knows and where the student may be having difficulties. It may also expand the teacher's evaluation of the student. Secondly, the student-teacher interaction helps the student develop reflective and self-evaluative skills The assessment process becomes a part of the learning process for both student and teacher. The richness of this dialogue and its benefits would be lost were traditional "objectivity" enforced and the teacher-student interaction prohibited.

In sum, using techniques oriented towards point estimation to warrant the quality of authentic assessments is out of place because the assumptions on which authentic assessments rest do not fit with traditional measurement theory. This means that it is inappropriate to use measurement theory to judge authentic assessments. The lack of fit between authentic assessments and the techniques used to judge traditional assessment approaches may explain why researchers and practitioners have had difficulty blending the two together. Clearly, a framework is needed for judging alternative assessments. This framework, though, needs to be congruent with the assumptions underlying alternative assessments.

The Constructivist Paradigm and its Assumptions

Guba and Lincoln's work developing the constructivist paradigm (Guba & Lincoln, 1989; Lincoln & Guba, 1985) is an important resource to those concerned about the quality of alternative assessments. Working from the framework of program evaluation, Guba and Lincoln questioned the value of the traditional, positivist research paradigm when evaluating social and educational programs. They rejected the positivist approach because they found a mismatch between the assumptions of the positivist paradigm and the real world circumstances confronting the evaluator. They proposed the constructivist paradigm as an alternative, believing its underlying assumptions offer a better fit for guiding program evaluation.

The constructivist paradigm is important to alternative assessments because of the synergy between the assumptions underlying the paradigm and alternative assessments. Their work provides a springboard for developing new ways of judging alternative assessments.

What are the Assumptions of the Constructivist Paradigm?

The constructivist paradigm is based on assumptions which are just the opposite of those which serve as the foundation for the positivist paradigm. The constructivist paradigm rejects the notion that reality is fixed and independent of the observer. Instead, this paradigm rests on the belief that reality is a construction made by an observer. (See Lincoln and Guba, 1985 for a discussion of possible ways that reality may be viewed.) Since each observer makes his or her own construction, there is no single construction of reality. Instead, there are multiple constructions of reality. Furthermore, "Constructions can be and usually are shared, ranging all the way from constructions about subatomic particles to those about cultural mores. That does not make them any more *real* but simply more *commonly assented to*" (Guba and Lincoln, 1989, p. 86, emphasis in the original). The researcher's job becomes one of adequately portraying and understanding the constructions (the realities) of the informants from whom information is collected.

The constructivist paradigm also differs from the positivist paradigm in its characterization the relationship between the inquirer and the object of inquiry. The constructivist approach follows logically from the assertion that there are multiple realities or constructions. Understanding an individual's construction requires a flexible data collection instrument, one capable of probing and clarifying that individual's point of view. The constructivist paradigm believes that the human researcher best fills this role. He/she has the kind of adaptability needed during data collection, ". . .the human can sense out salient factors, think of ways to follow up on them, and make continuous changes, all while actively engaged in the inquiry itself" (Lincoln and Guba, 1985, p. 107).

Relevance to Alternative Assessments

A reflection on authentic assessments and the constructivist paradigm suggests major parallels between these two areas. The first parallel is between the notion that achievement is multi-dimensional and the constructivist position on multiple realities. The second parallel is evident in how both authentic assessments and the constructivist paradigm stress the importance of interaction in the assessment and data gathering processes.

Lincoln and Guba developed the constructivist paradigm to guide program evaluation, not student assessment. Yet, the parallels between the two areas invite an exploration of how the constructivist paradigm might inform alternative assessment practices. No where is this exploration more needed than when focusing on the standards that should be used to judge the quality of alternative assessments since.

Validity, reliability, generalizability, and objectivity are criteria used to judge the quality of studies conducted within the positivist paradigm. Lincoln and Guba have rejected using these rigor criteria, as they refer to them, with the constructivist paradigm because of the incompatibilities of the assumptions underlying the two paradigms. The constructivist paradigm offers two sets of criteria, *trustworthiness* and *authenticity*, for judging studies.

The trustworthiness criteria parallel the rigor criteria and deal with the quality of a study. Lincoln (1990) described the development of the trustworthiness criteria in this way:

> The editors of education journals were interested in seeing naturalistic[1] case studies appear in print, but they were at a loss as to how to judge the rigor of those studies, and their reviewers were no better off. After his participation in a conference to deal with these issues, Egon took up the express intellectual task of the development of a set of criteria for judging the process of naturalistic inquiries. The result . . . was a set of criteria for judging whether or not any given inquiry was *methodologically* and *analytically* sound (emphasis in the original, pp. 70-71).

There are four elements to trustworthiness: credibility, transferability, dependability, and confirmability. They parallel the rigor criteria of validity, generalizability, reliability, and objectivity.

The second set of criteria Lincoln and Guba have proposed using to judge studies are authenticity criteria. These criteria deal with increasingly sophisticated levels of understanding and the increasing ability to take informed action. They were developed by considering the unique properties and concerns of the constructivist paradigm. As a result, they have no counterpart in the conventional research paradigm.

---

[1] Prior to the publication of *Fourth Generation Evaluation* in 1989, Lincoln and Guba referred to the constructivist paradigm as naturalistic inquiry. Guba's initial writing on trustworthiness was published in 1981.

Trustworthiness

Credibility

Credibility asks the researcher to establish that he/she has adequately represented the constructed realities of respondents. The application of this notion to the area of student assessment asks how confident is one of the accuracy of judgments, based on authentic assessments, about what a student knows and is able to do. Guba and Lincoln (1989) suggest several techniques that may be used to help establish credibility including: prolonged engagement, persistent observation, peer debriefing, progressive subjectivity, and member checks. The following is an overview of these techniques and the ways they might be applied to authentic assessments.

Prolonged engagement. This technique provides scope to the researcher's observations. It asks the researcher to be sufficiently immersed in the culture under study that he/she is able to both build trust and overcome distortions, "fronts," or misinformation. A goal is to become aware of the multiple factors impinging on subject under study. In an assessment situation, the technique of prolonged engagement suggests that the teacher develop rapport with students and an understanding of the factors that shape their performance. The day-to-day classroom seems to be an ideal setting for prolonged engagement to occur. The judgments that the teacher makes during the assessment process become grounded in on-going experiences with students.

Persistent observation. Persistent observation adds depth to the researcher's work and allows the researcher "to identify those characteristics and elements in the situation that are most relevant to the problem or issue being pursued and focusing on them in detail" (Lincoln & Guba, 1985, p. 304). In the area of student assessment, persistent observation means that the teacher is sufficiently engaged with the student to identify the salient factors which allow or prevent him/her to develop targeted skills and knowledge. The teacher may then use this information to build on strengths or diagnose and remediate weaknesses. In the classroom, prolonged and persistent observation complement each other.

A teacher knows that a student does not perform at a peak level every time assessment information is gathered. Furthermore, it is possible that the student may do poorly in some instances not because he/she lacks the needed knowledge or skills, but

because he/she does not understand what is being asked in the particular performance situation. Demonstrating that the judgments made about a student are credible requires that many performances or observations be made over time. Many observations are necessary because it is too easy for a judgment to be distorted if it is made on the basis of just one or two observations. Credible authentic assessments should be ongoing to allow a teacher's judgments to be grounded in experiences over time with the student.[2]

Peer debriefing. When using the technique of peer debriefing, the researcher reviews his/her findings, analyses, and conclusions with a disinterested peer. These reviews are conducted to give the researcher another perspective on his/her study, one which might support that work, challenge interpretations, or ask questions which lead to further insights into the subject matter.

In the context of assessment, peer debriefing could involve a teacher working with a principal or another teacher, reviewing assessment information about students. During the review, the teacher might present information about a student and receive information about whether his/her interpretations seem credible given the available information. Together they could discuss the next steps the teacher might take in the classroom when working with the student. Teachers using alternative assessments might work in pairs or small teams, periodically reviewing each other's work with students. A variant of this peer debriefing strategy could be used by instructors team teaching a class or who work with the same group of students. Teachers could periodically review assessment information about students and consider how their individual judgments corroborate each other.

Progressive subjectivity. No researcher goes into a setting with a blank mind, no preconceived ideas of the situation. Progressive subjectivity is designed to monitor the researcher's developing ideas, the goal being to ensure that the researcher's findings are not dictated by his/her own biases. This process may occur during peer debriefing or by reflecting on one's impressions and thoughts in a journal. A teacher could add to the credibility of authentic assessments by keeping a journal of his/her impressions of individual students and the classroom. This journal could be reviewed during the

---

[2] Shavelson and Baxter's (1992) findings from attempts to use performance assessments in science supports the need for many observations. They found that students must complete many performance tasks to give a reliable measure of their abilities in science .

debriefing process to show how the teacher's judgments were not shaped by the teacher's own biases.

Member checks. The constructivist researcher strives to understand the constructions of groups of respondents. Member checks entail reviewing the researcher's analyses and interpretations with members of those groups who provided information. This process helps ensure that the researcher has not distorted respondents' constructions.

Member checks in an assessment situation might involve discussions between a student and a teacher where they review the teacher's judgments of the student's work and compare them to the student's self-assessment. This process could provide valuable feedback about the accuracy of the teacher's judgments. A variant of this occurs in many classrooms where portfolios are used. When students select pieces to include in their portfolios, they complete a sheet explaining why they selected the piece. The sheet may also ask the child to reflect on such areas as the quality of the piece or why it is important to them. By reviewing a student's comments, the teacher may obtain feedback about his/her own judgments of the student.

Some schools expand the dialogue between the student and teacher by including the student's parents. Parents add to student portfolios their own reflections on the student's work. Parent comments may support the observations that the teacher and student have already made. Alternatively, a parent might have conflicting information or perspectives which would modify interpretations that a teacher or student has already made. Including the parent in the student-teacher dialogue provides further credibility to the teacher's judgments. Their inclusion also helps to strengthen the home-school link.

Transferability

Under the traditional research model, the researcher seeks to develop generalizations about the way the world is based on his/her data collection and analyses. Ideally, these generalizations take the form of law-like statements. The ability to make generalizations comes out of the view that reality is fixed and apart from the observer.

The constructivist paradigm, in rejecting the notion of a reality independent of the observer, also dismisses the possibility of developing generalizations (Lincoln & Guba,

1985). The paradigm replaces the concept of generalizations with that of working hypotheses. While gathering information, the constructivist researcher develops working hypotheses about how constructs fit together in the contexts where data were gathered. These working hypotheses are open to revision as data gathering proceeds. Since data collection is context bound, the researcher's working hypotheses are too. It is therefore incumbent upon the researcher to specify the context(s) in which those working hypotheses were developed (the sending context). This allows others to judge the contexts in which the findings will be transferable (receiving contexts). Transferability increases the more that the sending and receiving contexts are alike.

Thick Description. Transferability is established through the technique of thick description which involves providing extensive information about the sending context. This information facilitates judgments about the overlap and match between the sending and receiving contexts. Applied to assessment, transferability asks to what extent judgments made about knowledge and skills displayed in one situation are applicable to other situations. Judgments could be made about one or a group of students. Teachers may facilitate these judgments by providing full information about performance tasks and the contexts to which they were completed. Then inferences could be made about whether the skills displayed in one setting are relevant to or likely to be displayed in another setting.

Wiggins (1989) noted that authentic assessments are valued because their performance is valued in the real world. The issue of transferability highlights the importance of carefully considering the purpose of any authentic assessment. If the results from an assessment are to be truly valuable, educators, when designing tasks, must keep in mind the receiving context where the performance is valued. Ideally, teachers should use performance tasks which have broad transferability. Otherwise, students may be assessed in areas which have little value beyond one or two contexts.

Dependability and Confirmability

Dependability and confirmability are the last two trustworthiness criteria and parallel, respectively, the traditional concepts of reliability and objectively. Dependability issues relate to questions of whether the inquiry process is documentable and consistent with good research practices. Dependability is established by making the inquiry process publicly inspectable. Any shifts or changes in the process " . . . need to be both tracked and trackable (publicly inspectable), so that outside reviewers can explore the process, judge the decisions that were made, and understand what salient factors in the context led the evaluator to the decisions and interpretations made" (Guba & Lincoln, 1989, p. 242). Applied to student assessment, dependability rests on being able to show the process used to make decisions about a student's skills and abilities may be documented and tracked

Confirmability involves demonstrating that the information collected during a study and the ways it is interpreted are not a function of the researcher's biases. The conventional approach to research addresses this issue through the use of a methodology which assures that the researcher does not bias the data. The constructivist paradigm shifts the focus from "objective," a characteristic of the researcher, to "confirmable," a quality of the data themselves. "This means that the data (constructions, assertions, facts, and so on) can be tracked to their sources, and that the logic used to assemble the interpretations into structurally coherent and corroborating wholes is both explicit and implicit in the narrative of a case study" (Guba & Lincoln, 1989, p. 243).

Confirmability and Dependability Audit. Guba and Lincoln (1989; Lincoln & Guba, 1985) suggest the primary way to establish confirmability and dependability is through an audit process. The auditing metaphor comes directly from the fiscal audit that is conducted to attest to the financial condition of a company. The fiscal auditor examines the accounting process of the company and its financial data and then makes a statement about whether the financial statement provides a fair and accurate representation of the company's financial condition. In a similar fashion, audits can be conducted of an inquiry. A dependability audit focuses on the process by which data were collected. The confirmability audit asks whether data can be tracked to their

original sources and if the inferences made during the study and its outcomes are rooted in these data. Both of these audits may be conducted simultaneously.[3]

Authentic assessments have processes and products that may be audited. It is possible to identify many elements in an assessment system that are process related. These elements include such things as: the way that a performance assessment, a portfolio system, performance standards, or scoring rubrics are developed; the ways that standards or rubrics are applied; the processes students use when deciding what to put in a portfolio; the factors a teacher considers when making judgments about a student's performance or assigning a grade; and the process for training raters who will score a product holistically.

Understanding the importance of the audit function contributes significantly to developing and implementing superior systems of authentic assessment. A traceable record, an audit trail, documenting how elements of the system were developed or operate may be constructed from the outset. This trail could include such things as minutes of meetings where assessment systems, standards, or scoring rubrics were developed. The minutes would ideally show the various factors that were considered in the development process. A holistic scoring session might be videotaped to show how raters learned to apply scoring anchors or the training procedures might be described in writing. The key point here is once an audit trail is recognized as critical, a range of information may be incorporated into it.

The primary products in authentic assessments are performances and scores. Here both performances and scores are broadly considered. A performance could be: a piece of student writing in a portfolio; the portfolio itself containing many pieces of a student's writing; a student's log from a hand-on performance task in science; or a student exhibition of a senior research project. Scores could encompass: a teacher's rating of a student's assignment, project, or portfolio; the grade the teacher assigns; the rating given to a performance task or writing sample; a student's own self-assessment of his/her work.

A performance may be easy to review or refer to during an audit because it most likely is physical--a piece of paper or perhaps a videotape. A score is more difficult to

---

[3] Schwandt & Halpern (1988) and Lincoln & Guba (1985) provide more complete information about auditing program evaluations.

audit because scoring involves making a judgment. This is another reason why it is important to consider auditing when developing an assessment system. For example, a teacher, in commenting about a student's work, might say the piece is better than a past draft or an earlier piece of writing, "Your work is improving." Such a statement could be difficult to audit because the teacher has not indicated the evidence on which her judgment was based. A better statement might be, "Your work is improving because you're now beginning paragraphs with a topic sentence." This second statement allows an auditor to see why the teacher made her judgment. Providing the more auditable feedback also has the benefit of letting the student know why the teacher believes the writing is improving and may serve as the basis for a member check.

Consider a second example. When working with portfolios, students are often asked to choose what they consider a best piece of work or a piece that shows improvement. Asking the student to comment on why he/she made the particular choice provides an auditable record of the child's perspective. This strategy also lets the teacher understand the child's reasoning which, in some cases, may dramatically shift the teacher's perception of the child. It may only take a few words or phrases for a teacher or student to explain their evaluation of a piece of work. A brief phrase, though, may add significantly to the auditability of an authentic assessment system.

The Pittsburgh schools have begun to experiment with auditing student portfolios (LeMahieu, Eresh, & Wallace, 1992). The district convened a team of 20 independent, external auditors who were asked, "to inspect the procedures, methods, and logistics of the portfolio assessment carried out the district so as to satisfy itself that the results of the application of those assessment procedures would represent a fair and critically honest accounting in good faith describing the performance of the school district" (LeMahieu, Eresh & Wallace, p. 7). The Pittsburgh experience confirmed that audits are possible and may have a positive effect on teaching and learning.

Authenticity and Fairness

The trustworthiness criteria represent one way of judging the quality of constructivist inquiries. They were developed to parallel the rigor criteria used to judge research conducted using the positivist paradigm. Authenticity and fairness, in contrast,

were developed by questioning how to judge the quality of a constructivist study if one considered only the attributes indigenous to constructivism (Lincoln, 1990).

A constructivist evaluation seeks to faithfully represent the constructions of program stakeholders and participants. The several forms of authenticity identified by Guba and Lincoln (1986; 1989; Lincoln, 1990) emphasize greater awareness of the constructions one holds (ontological authenticity), greater awareness of the constructions of others (educative authenticity), the call to action resulting from a study (catalytic authenticity), and the ability to act for oneself or one's group in the political arena, empowerment (tactical authenticity). "In conventional inquiry, pure process leads to pure results. In constructivist inquiry, process is only one means of determining the utility, responsibility, and fidelity of the inquiry. Action and understanding were other components of the judgments regarding the goodness of any inquiry" (Lincoln, 1990, p. 72).

Fairness is an over-riding concern in constructivist inquiry. The fairness criterion demands a balanced presentation of all constructions and values. Additionally, groups should negotiate meanings and actions from equal positions of strength, with no group being disadvantaged as a result of the inquiry (Guba & Lincoln, 1986).

The authenticity criteria with their focus on knowing, action, and fairness, blend well with authentic assessments. The concepts of ontological, educative, catalytic, and tactical authenticity speak to the heart of the education process. These assessments provide information which deepens the teacher's and the student's understanding of the student's knowledge and abilities. The authenticity criteria also serve as a reminder of the importance of the teacher-student interaction to this understanding. Additionally, the assessment process helps focus future teaching and learning activities.

Authenticity is primarily demonstrated in two ways (Guba and Lincoln, 1989). First, there is stakeholder testimony. Members of stakeholder groups may be questioned about how each of the authenticity criteria was honored during a constructivist inquiry. Second, the audit trail may be reviewed for evidence of fairness and authenticity. Parallel processes could be used with authentic assessments. Interested stakeholders such as students, teachers, and parents could be interviewed to determine how authentic assessments were fair, facilitated deeper understanding, and empowering. The audit trail

could also be reviewed for documentation showing how authenticity and fairness were honored.

Messick (1989) and Linn, Baker, and Dunbar (1991) have discussed the consequential validity of assessment pointing to the importance of considering the consequences of assessment. By consequences, these authors are referring to actions which may occur prior to assessment and have differential impact on students' performance. Such actions would include over-emphasizing content areas likely to be included on a test or assessment while under-emphasizing of other areas.

Lincoln and Guba's use of fairness and authenticity, particularly catalytic and tactical authenticity, suggest that it is also important to consider the consequences that result *after* assessment occurs. Catalytic authenticity asks whether action is taken based on the assessment (inquiry) process. Tactical authenticity is an empowerment criterion. Fairness asks that no group or stakeholder be disadvantaged because of their participation in the assessment (inquiry) process. Combining these ideas together suggests that an assessment lacks authenticity when there are negative consequences which occur based on the outcomes of the assessment. This is a broader notion of consequences than evidenced in most discussions of validity (Moss, 1992).

One implication of this broader concern with consequences is that the developers of assessment systems must pay attention to what happens to students after they are assessed. Negative post-assessment consequences undermine the authenticity and quality of an assessment system and should be taken into account when the system is developed. Clearly, developers will need to identify what constitutes a negative, post-assessment consequence. Would an assessment system lack authenticity if afterwards, students were tracked or some students were not promoted to the next grade? Such questions have policy implications and merit consideration as assessment systems are designed regardless of whether the systems are based on standardized tests or authentic assessments.

Conclusion

The movement toward authentic assessments promises more valuable approaches to informing the teaching-learning process than the testing approach which has dominated education. As new assessment systems are developed, concerns about their quality are

justified and widely shared. However, attempts to address these concerns by relying on traditional standards like reliability, validity, generalizability, and objectivity will prove unsatisfying to those designing these new systems. This dissatisfaction is caused by the failure to recognize that the traditional standards and new assessment approaches grow out of different assumptions.

Lincoln and Guba's work with the constructivist paradigm offers new lenses for viewing alternative assessments and judging their quality. This paper discussed how two sets of criteria for judging constructivist inquiries, trustworthiness and authenticity, might be applied to authentic assessments. Many of the elements needed to judge authentic assessments using these criteria are already in place in some alternative assessment systems. Other elements could be readily added.

References

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. Educational Communication and Technology Journal, 29, 75-92.

Guba, E. G., & Lincoln, Y. S. (1989). Fourth Generation Evaluation. Newbury Park, CA: Sage.

LeMahieu, P. G., Eresh, J. T., & Wallace, R. C., Jr. (1992). Using student portfolios for public accounting. The School Administrator, (December, 1992), 8-15.

Lincoln, Y. S. (1990). The making of a constructivist: A remembrance of transformations past. In E. G. Guba (Ed.), The Paradigm Dialog (pp. 67-87). Newbury Park, CA: Sage.

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic Inquiry. Newbury Park: Sage.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. Educational Researcher, 20(8), 15-21.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. Educational Researcher, 18(2), 5-11.

Moss, P. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. A paper presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, San Francisco, CA.

Schwandt, T. A., & Halpern, E. S. (1988). Linking auditing and metaevaluation: Enhancing quality in applied research. Newbury Park, CA: Sage.

Shavelson, R. J. & Baxter, G. P. (1992) What we've learned about assessing hands-on science. Educational Leadership, 49(8), 20-25.

Wiggins, G. (May, 1989). A true test: Toward more authentic and equitable assessment. Phi Delta Kappan, 703-713.

Wolf, D. P., Bixby, J., Glen, J., III, & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed.), Review of Educational Research (pp. 31-74). Washington, D. C.: American Educational Research Association.